The Akrivia Health
Record-Level Data
Access Service

1. The Akrivia Health Dataset

Akrivia Health (Akrivia) curates a dataset of over 5.1 million patients' electronic health records (EHRs) from secondary care psychiatric healthcare organisations (HCOs) in England and Wales, available for research access.

- In partnership with the NHS, Akrivia has developed a proven data governance, security and access model to overcome the barriers to research application of EHR data.

- Akrivia now provides this free service to a network of 20 secondary care psychiatric HCOs in England and Wales, with a paired secure data access service for a growing network of academic and industry research users.

- At the time of writing, the Akrivia dataset includes 5.1 million+ anonymised longitudinal EHR records, available for research application by third parties (for example, research teams from academia, industry, government and the third sector).

- This dataset includes patients with severe depression, anxiety and eating disorders, schizophrenia, and dementia. This also includes comprehensive chronologies of care for admissions to inpatient wards, and outpatient treatments in community settings by HCOs in the Akrivia network.

2. Data Enrichment

As part of our curation service on behalf of HCOs, Akrivia harmonises data, incorporates publicly available data (Index of Multiple Deprivation), actively facilitates linkages to other datasets (primary care) and extracts key clinical features (medication, diagnosis, signs & symptoms etc) from free-text clinical notes via an AI-enabled data structuring pipeline.

- Akrivia enriches HCO data with a bespoke data structuring pipeline to address the high proportion of critical clinical information captured in free-text fields (e.g. clinical notes, referral letters etc.) in NHS psychiatric EHRs. We then apply a range of AI and rule-based natural language processing (NLP) models to extract research-relevant features from unstructured fields, creating uniquely deep structured data on patients' care and clinical state including (but not limited to) the following clinical features:

  I.   **Medication** – presently, NHS secondary care psychiatric services only record medication information in unstructured text fields. Akrivia's data structuring pipeline is therefore critical for mapping treatment pathways, capturing response to treatment, identifying treatment resistance etc.

  II. **Diagnosis** – although psychiatric EHRs do include structured diagnosis fields, these are often underutilised (completion rates range from ~60% to as low as ~5% across HCOs in the Akrivia network). Diagnosis information is also recorded in unstructured fields, and consequently Akrivia's data structuring pipeline provides more complete (and often earlier) diagnosis information than structured data, as well as capture of negated or potential diagnoses, and indications of family diagnostic history.

  III. **Signs and Symptoms** – psychiatric clinical notes include rich descriptions of patients' symptoms, with a level of detail typically absent from other administrative data sources. Through Akrivia's data structuring pipeline, these symptom descriptions can be translated into a structured format, allowing for deep, transdiagnostic profiling, subtyping, risk prediction, and analysis treatment response/adverse effects.

The current Signs & Symptoms models extract, classify and map 420 distinct symptoms (mapped to SNOMED IDs) across the full Akrivia database.

- In addition to the data structuring pipeline, Akrivia also carries out extensive harmonisation and data linkage to further enrich the EHR dataset. Structured fields from the source EHR (e.g., demographics, referrals, appointments etc) are harmonised across HCOs to facilitate their use in research. Before being redacted, post-codes are linked to LSOA codes (a higher-level geographic indicator), allowing linkage to a range of public data sources such as the Index of Multiple Deprivation. Akrivia has also established a data linkage collaboration with OPCRD, who provide a route to access linked primary care records for patients in the Akrivia database.

3. Record-Level Secure Data Access

The Akrivia database can be accessed via a paid secure data access service, allowing researchers to safely interrogate an anonymised (but still record-level) version of the full, network-wide dataset within Akrivia's Secure Data Environment (SDE).

- The Akrivia Health Record-Level Data Access Service (RLDA) thus provides researchers with secure access to an anonymised, record-level version of this dataset, hosted within Akrivia's Secure Data Environment (SDE).

This is a paid service, in order to ensure its sustainability, and continued development of the dataset. Non-commercial users (academia, government, charities) receive highly discounted rates, with a further low-cost option for Student Researchers. Cost is based on duration of access and number of users, as follows:

RLDA Licence	Commercial From £75,000	Non-Profit From £15,000	Student £5,000
Six months' data access	x	x	*3 months*
User-defined EHR dataset	x	x	x
Up to 3 user licences	x	x	*1 user licence*
Study setup & management	x	x	x
Access to online training + knowledge base	x	x	x
Data analytics support (up to 24 hours)	x	x	-
Technical support	x	x	x
Data archiving	x	x	x
Akrivia TRE workspace licence	x	x	x
Standard Spec VM	x	x	x
Named account manager	x	x	-

- To arrange access, please contact [support@akriviahealth.com.](mailto:support@akriviahealth.com) The access process is as follows:

  I. Akrivia will establish a master services agreement (MSA) with the applicant's host institution (if an MSA with that organisation does not exist already).

  II. While the MSA is being agreed, applicants can prepare a project application for submission to Akrivia's Data Access Committee (DAC). The DAC includes representation from public and patient participants, governance specialists and scientists. It is responsible for ensuring approved access is proportional, has clearly defined research objectives and outputs, and is oriented towards patient benefit. DAC will review all applications within 3 weeks of submission.

  III. Once the DAC application is approved, designated researchers must sign an honorary research agreement with Akrivia, permitting access to the anonymised data under the terms of our service agreement with our NHS data partners. These agreements are required under the terms of our service agreement with the NHS. They do not limit researchers' ability to publish or utilise work carried out during the project.

- Once the MSA, DAC application and HRA are signed, payment can be processed, and access provided. Access is provided via a virtual workspace (hosted by Amazon Web Services) created for the project.

- A project-specific portion of the anonymised Akrivia database, (defined as part of the DAC application) will be placed in the project workspace, alongside standard analytic software (R, Python) with locally-hosted package repositories.

- Data is provided in an anonymised, record-level format within the project workspace. Aggregate statistics, tables and figures can be exported through a file-out process managed by Akrivia's Governance team.